**Big Data Final Project**  
Team Members:  
Amit Nair  
Manish Reddy Chinthalaphani  
Vajrender Kumar Mangalapally

This final project contains multiple analytics. The analytics are described below

**Analytic 1:** Analyzed the taxi and uber usage based on the climatic condition

Things to be considered for this analytic are the date at which the ride was taken place and the weather corresponding to that day. Joined both the taxi data table and the weather data table based on the date. We used the PIG script to clean and join the data for this analytic. Visualized the trips based on the climatic condition using Tableau.

**Analytic 2:** Analyzed subway, taxi and uber data to count the number of rides which are taking place near the subway and predicting the names of the top 5 subways for pick-ups.

In this analytic, the locations of all the subway stations present, pickup locations of the taxi and uber data are taken into consideration. Here the location corresponds to the latitude and longitude. Based on the number of rides taking place near the subways, we ranked the top 5 subway stations for pick up. PIG script used to clean the data and Hive queries used hive queries used to join subway data, taxi data / uber data.

**Analytic 3:** Suggesting the driver where to go to get the next ride after a drop off has taken place.

In this analytic, based on a particular drop off location we are calculating the chances of getting a ride for the driver within the 2 miles, 4 miles and 6 miles radius. Here the drop-off latitude and longitude are user defined, in our analytic we have taken a random location and calculated the chances of getting a ride for the driver within a particular radius at a particular time. For this analytic, we used Hive queries to calculate the rides present with 2 miles, 2 to 4 miles, 4 to 6 miles radius separately and put them all together into a table and shown it visually on Tableau.

**Analytic 4:** Calculating the revenue earned for each day, and the number of rides taken place

In this analytic, we calculated the revenue earned per day and the number of trips that have taken place, based on the climatic condition, here we filtered the data based on the climatic condition of Snow and shown if there are any drop in the revenue. Here we visualized the data on Tableau to show the earning for each day in the past 2 years i.e. 2014 and 2015 and observed the trends of income earned on weekdays and weekends. Used Hive query to calculate and store the data.

**Analytic 5:** Comparison of usage patterns of Yellow Taxi and Uber.

In this analytic, we are comparing the usage of both Yellow taxi and Uber based on the number of pick-ups taken by both of these taxi services. Hive script used to run the analytic and visualized this usage pattern for every month and shown it on Tableau.